SRC-TR-87-91

Arbitrarily Varying Channels with Constrained Inputs and States

by

Imre Csiszar
Prakash Narayan

To appear, IEEE Transactions on Information Theory

ARBITRARILY VARYING CHANNELS WITH CONSTRAINED INPUTS AND STATES*

Imre Csiszár and Prakash Narayan

ABSTRACT

Random coding theorems are proved for *discrete memoryless arbitrarily varying channels* (AVCs) with constraints on the transmitted codewords and channel state sequences. We consider two types of constraints: *peak* (i.e., required for each n-length sequence almost surely) and *average* (over the message set or over an ensemble). For peak constraints on the codewords and on the channel state sequences, the AVC is shown to have a (strong) random coding capacity. If the codewords and/or the channel state sequences are constrained in the average sense, the AVCs do not possess (strong) capacities; only ϵ-capacities are shown to exist.

* This research was sponsored by the Systems Research Center at the University of Maryland under NSF grant No. OIR-85-00108 and by the Minta Martin Fund for Aerospace Research from the University of Maryland. I. Csiszár is with the Mathematical Institute of the Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary. P. Narayan is with the Electrical Engineering Department, University of Maryland, College Park, MD 20742, USA.

1. Introduction

A discrete memoryless *arbitrarily varying channel* (AVC) is a model for a communication channel with unknown parameters that may vary with time in an arbitrary and unknown manner during the transmission of a codeword. The encoder transmits over the channel, once in each unit of time i, a symbol x_i from a *finite* alphabet $\mathcal{X}$. The transmitted symbol is received at the output of the channel as a symbol y_i taking values in a *finite* alphabet $\mathcal{Y}$. The use of the channel through n units of time, i.e., "n uses of the channel" can be modeled by a stochastic matrix $W^n : \mathcal{X}^n \to \mathcal{Y}^n$, where $W^n(\mathbf{y} \mid \mathbf{x}, \mathbf{s})$ is the probability that a transmitted sequence $\mathbf{x} = (x_1, ..., x_n)$ is received as the sequence $\mathbf{y} = (y_1, ..., y_n)$ given that the channel resided in the sequence of *states* $\mathbf{s} = (s_1, ..., s_n)$. Here, the state s_i, at each time unit i, belongs to a *finite* set $\mathcal{S}$ of states, and may vary with i in an arbitrary manner. The transmitter and receiver strive to construct codes for reliably transmitting information across such a channel.

There is a large variety of coding problems for the AVC, depending on the nature of the error criteria used (*average* or *maximum* error), on the permissible coding strategies (*correlated randomization* in *encoding* and *decoding, randomization* in *encoding* only, or *no randomization*), and on whether or not the codeword and state sequences are selected with a *knowledge of each other.*

Since the introduction of the AVC by Blackwell, Breiman and Thomasian [11], considerable progress has been made in the study of these problems. Much of the work is summarized in Csiszár and Körner [12, Chapter 2, Section 6] (see also Wolfowitz [24]); we cite only a few results here. The pioneering work of Blackwell, Breiman and Thomasian [11] used *random* codes, that is, the encoder and decoder were chosen by a random experiment whose outcome had to be available to *both* the encoder and the decoder. The evident practical drawbacks of such a scheme led to a study of *deterministic* codes for AVCs [22] with a *maximal* error probability criterion. Ahlswede and Wolfowitz [7] determined the corresponding capacity for AVCs with a binary output alphabet. For general outputs, the problem is still unsolved and includes Shannon's famous zero-error capacity problem [2, 23] as a special case. In a major breakthrough, Ahlswede [5] determined the capacity of a fairly large class of AVCs for the maximal probability of error criterion. The best results yet on this problem are due to Csiszár and Körner [13]. For the *average* probability of error criterion, the basic AVC coding theorem is due to Ahlswede [4] who proved that the capacity for deterministic codes, if positive, is

always equal to that for random codes. However, the random coding capacity may be positive when the deterministic capacity is zero; a necessary and sufficient condition for the positivity of the latter will be given in Csiszár and Narayan [14].

All the results mentioned above are for the case when the transmitted and state sequences are chosen without any knowledge of each other. Since this case is being considered in the present paper, from among results for other cases, we only cite a remarkable recent paper of Ahlswede [6]. In [6], using previous results of Gel'fand and Pinsker [18], the capacity problem is completely solved for the case when the state sequence is known to the encoder.

Continuous alphabet AVCs are less understood than their discrete counterparts, and all the available results refer to the Gaussian case [3, 9, 10, 20]. Of particular interest to us, and in fact a strong motivation for the present paper, were the *Gaussian* AVCs of [20]. A Gaussian AVC (GAVC) in the sense of [20] is a discrete-time, memoryless Gaussian channel with *input power constraint* P_T and noise power N_e, further corrupted by an additive "jamming signal" whose statistics may be arbitrary and unknown, subject only to a (known) *power constraint* P_J. Considering two types of power constraints, viz., *peak* and *average*, it was shown for peak power constraints on *both* the input (i.e., codeword) and jamming sequences that the GAVC had a random coding strong capacity. For the remaining combinations of peak and average power constraints on the input and jamming sequences, the GAVCs were shown in [20] not to possess strong capacities.

This paper considers problems analogous to those in [20] for a general class of *discrete* AVCs with peak and average constraints (defined in Section 2) on the input and state sequences. Preliminary results are available in [19]. As in [20], it turns out that the random coding strong capacity exists only in the case of peak constraints on both the input and state sequences, while otherwise the ϵ-capacities do depend on ϵ. This is explained by the fact that AVCs with average state constraints are similar to ordinary "averaged channels," for which a strong capacity does not exist (cf. Ahlswede [1]), while under average input constraints not even a discrete memoryless channel has a strong capacity.

The capacity problem for the AVC under constraints using deterministic codes will be addressed in a forthcoming paper [14]. Here we only mention that the proof technique of Ahlswede [4] may not work in the constrained case and, in fact, the deterministic average error capacity may be *positive and strictly less than the random code capacity.*

In the remainder of this paper, we introduce the terminology and definitions in Section 2, and prove our results in Section 3. Section 4 is devoted to a discussion of these results.

2. Terminology and Definitions

We have adopted much of our terminology and definitions from [12].

In particular, $\mathcal{X}, \mathcal{Y}$, and $\mathcal{S}$ denote *finite* sets, and X, Y, and S random variables taking values in these sets. The distributions (resp. joint distributions) of such random variables are denoted by P_X, P_Y, P_{XY}, etc., while conditional distributions are denoted by $P_{Y|X}, P_{Y|XS}$, etc. A *channel* $W : \mathcal{X} \to \mathcal{Y}$ is given by a transition probability matrix $\{W(y \mid x) : x \epsilon \mathcal{X}, y \epsilon \mathcal{Y}\}$. A *discrete memoryless channel* (DMC) $\{W\}$ defined by W is a sequence of channels $\{W^n : \mathcal{X}^n \to \mathcal{Y}^n\}$ where

$$W^n(\mathbf{y} \mid \mathbf{x}) \triangleq \prod_{i=1}^{n} W(y_i \mid x_i), \tag{2.1}$$

with $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$.

Let $\mathcal{W} = \{W(\cdot \mid \cdot, s), s \epsilon \mathcal{S}\}$ be a family of channels $W : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ represent the input and output alphabets; $s \epsilon \mathcal{S}$ denotes the *state* of the channel and can be interpreted as an index identifying a particular $W \epsilon \mathcal{W}$. For n-lengths sequences, the transition probabilities corresponding to a sequence of states $\mathbf{s} = (s_1, ..., s_n)$ are assumed to be given by

$$W^n(\mathbf{y} \mid \mathbf{x}, \mathbf{s}) \triangleq \prod_{i=1}^{n} W(y_i \mid x_i, s_i). \tag{2.2}$$

The family of channels $W^n(\cdot \mid \cdot, \mathbf{s}) : \mathcal{X}^n \to \mathcal{Y}^n$, $\mathbf{s} \epsilon \mathcal{S}^n$ will be denoted by $\mathcal{W}^n$.

A (discrete memoryless) *arbitrarily varying channel* (AVC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and set of states $\mathcal{S}$ is a sequence $\{\mathcal{W}^n\}_{n=1}^{\infty}$ as above; henceforth, we shall denote it simply by $\{\mathcal{W}\}$.

A *code of block-length n* is a pair of mappings $f : \mathcal{M} \to \mathcal{X}^n, \phi : \mathcal{Y}^n \to \mathcal{M}$, and its *rate* is $\frac{1}{n} \log \mid \mathcal{M} \mid$, where $\mid \mathcal{M} \mid$ denotes the cardinality of the *message set* $\mathcal{M}$. The performance of the code (f, ϕ) on any channel $W^{(n)} : \mathcal{X}^n \to \mathcal{Y}^n$ is evaluated in terms of its rate and the decoding error probabilities. The *probability of error* for the message $m \epsilon \mathcal{M}$ is given by

$$e_m = e_m(W^{(n)}, f, \phi) \triangleq 1 - W^{(n)}(\phi^{-1}(m) \mid f(m)). \tag{2.3}$$

The corresponding *average error probability* is

$$\bar{e} = \bar{e}(W^{(n)}, f, \phi) \triangleq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m. \tag{2.4}$$

In particular, for $W^{(n)} = W^n(\cdot \mid \cdot, \mathbf{s})$, the dependence of the error probabilities on the state sequence $\mathbf{s} \in \mathcal{S}^n$ will be indicated by writing

$$e_m(\mathbf{s}) = e_m(\mathbf{s}, f, \phi) \triangleq e_m(W^n(\cdot \mid \cdot, \mathbf{s}), f, \phi); \tag{2.5}$$

$$\bar{e}(\mathbf{s}) = \bar{e}(\mathbf{s}, f, \phi) \triangleq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m(\mathbf{s}). \tag{2.6}$$

A *random code* (F, Φ) is a random variable taking values in the family of all codes (f, ϕ) with the same blocklength n and the same message set $\mathcal{M}$.

We now impose constraints on the input (transmitted codeword) sequences and define random codes that satisfy these constraints. Let g be a nonnegative-valued function on $\mathcal{X}$, and let

$$g(\mathbf{x}) \triangleq \frac{1}{n} \sum_{i=1}^{n} g(x_i) \tag{2.7}$$

for $\mathbf{x} = (x_1, ..., x_n)$ in $\mathcal{X}^n$. A random code (F, Φ) is said to satisfy a *peak input constraint* Γ, if for all $m \in \mathcal{M}$,

$$g(F(m)) \leq \Gamma \text{ almost surely (a.s.).}$$

It satisfies a *message average (m-average) input constraint* m-Γ, if $\bar{g}(F) \leq \Gamma$ a.s., where

$$\bar{g}(F) \triangleq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} g(F(m)), \tag{2.8}$$

and a *code-ensemble/message average (cm-average) input constraint cm-F*, if $E\overline{g}(F) \leq \Gamma$. Clearly, the peak input constraint Γ is stronger than the m-average input constraint m-Γ which, in turn, is stronger than the cm-average input constraint cm-Γ.

Remarks: Even what we term a peak input constraint is in a sense, an average constraint, as the constraining function g is defined by averaging over n time units (cf. (2.7)). We feel that the term "peak" is justified in comparison with the other types of constraints, and will not lead to ambiguity. It would also be possible to consider a fourth type of input constraint, namely

$$Eg(F(m)) \leq \Gamma \text{ for all } m \epsilon \mathcal{M}.$$

This, however, would not lead to a new problem because for any random code (F, Φ) there exists another random code (F', Φ') with the same message set such that for every $m \epsilon \mathcal{M}$ and every channel $W^{(n)} : \mathcal{X}^n \to \mathcal{Y}^n$,

$$Eg(F'(m)) = E\overline{g}(F), \; Ee_m(W^{(n)}, F', \Phi') = E\overline{e}(W^{(n)}, F, \Phi). \tag{2.9}$$

To obtain this (F', Φ') we may suppose that $\mathcal{M} = \{1, ..., M\}$. Then, denoting by Z a random variable independent of (F, Φ) and uniformly distributed over $\mathcal{M}$, we set

$$F'(m) \triangleq F(m + Z \; (\text{mod} M)); \quad \Phi'(\mathbf{y}) = \Phi(\mathbf{y}) - Z \; (\text{mod} M).$$

This fact also shows that for random codes it does not matter whether we adopt the average or maximum probability of error performance criterion.

Constraints can also be imposed on the sequence of channel states as follows. Let ℓ be a nonnegative-valued function on $\mathcal{S}$, and let

$$\ell(\mathbf{s}) \triangleq \frac{1}{n} \sum_{i=1}^{n} \ell(s_i), \tag{2.10}$$

for $\mathbf{s} = (s_1, ..., s_n)$ in $\mathcal{S}^n$. We also consider *random* state sequences $\mathbf{S} = (S_1, ..., S_n)$. Throughout this paper, it will be *assumed that the transmitted and state sequences are chosen without*

any knowledge of each other. Mathematically, this is reflected by the assumption that *the random variables* (F, Φ) *and* S *are statistically independent.* We say that S satisfies a *peak state constraint* Λ, if

$$\ell(\mathbf{S}) \leq \Lambda \text{ a.s.},$$

and satisfies an *average state constraint* a-Λ if

$$E\ell(\mathbf{S}) \leq \Lambda.$$

Clearly, the latter constraint is the weaker one.

For convenience, we shall assume that

$$\min_{x \epsilon X} g(x) = \min_{s \epsilon S} \ell(s) = 0,$$

and all input and state constraints will be considered with $\Gamma > 0, \Lambda > 0$.

Definition 2.1 For $0 < \epsilon < 1$, $\Lambda > 0$, an (n, Λ, ϵ)-*random code* for the AVC $\{\mathcal{W}\}$ is a random code (F, Φ) of blocklength n satisfying

$$E\bar{e}(\mathbf{s}, F, \Phi) \leq \epsilon \text{ for } \mathbf{s} \text{ in} S^n \text{ with } \ell(\mathbf{s}) \leq \Lambda. \tag{2.11}$$

Further, an $(n, a\text{-}\Lambda, \epsilon)$-random code for the AVC $\{\mathcal{W}\}$ is a random code (F, Φ) satisfying

$$E\bar{e}(\mathbf{S}, F, \Phi) \leq \epsilon \tag{2.12}$$

for all random state sequences $\mathbf{S} = (S_1, ..., S_n)$ meeting the a-Λ constraint $E\ell(\mathbf{S}) \leq \Lambda$. Clearly, every $(n, a\text{-}\Lambda, \epsilon)$-code is also an (n, Λ, ϵ)-code.

Definition 2.2 Given $0 < \epsilon < 1$, a non-negative number R is an ϵ-*achievable rate* on the AVC $\{\mathcal{W}\}$ under peak (resp. m-average resp. cm-average) input constraint Γ (resp. m-Γ, resp. cm-Γ) and peak state constraint Λ if for every $\delta > 0$ and every sufficiently large n there exist (n, Λ, ϵ)-random codes with rates $\geq R - \delta$ and satisfying the corresponding input constraint. The ϵ-achievable rates under peak (resp. m-average or cm-average) input constraint and

average state constraint are defined similarly but with $(n, a\text{-}\Lambda, \epsilon)$ random codes. Finally, R is an *achievable rate* under any pair of input and state constraints if it is ϵ-achievable for every $0 < \epsilon < 1$.

Definition 2.3 The maximum of all ϵ-achievable rates under a pair of input and state constraints is called the (random coding) ϵ-*capacity* of the AVC under these constraints. If it does not depend on ϵ, its value is called the *strong capacity*. Otherwise, the limit of the ϵ-capacity as $\epsilon \to 0$ or, equivalently, the maximum of all achievable rates, is called the *(weak) capacity*.

The ϵ-capacity under input constraint A and state constraint B will be denoted by $C_\epsilon(A, B)$, where A stands for either of Γ (peak) or m-Γ (m-average) or cm-Γ (cm-average) and B stands for Λ (peak) or a-Λ (average).

3. Random Coding Theorems

Our main results are random coding theorems determining the ϵ-capacities $C_\epsilon(\Gamma,\Lambda)$, $C_\epsilon(\Gamma, a\text{-}\Lambda)$, $C_\epsilon(m\text{-}\Gamma,\Lambda)$, and $C_\epsilon(cm\text{-}\Gamma,\Lambda)$ of the AVC $\{\mathcal{W}\}$ under either of the three kinds of input constraints with peak state constraint, and under the peak input constraint with average state constraint. First we introduce some notation and prove two technical lemmas.

Given an AVC $\{\mathcal{W}\}$ and a random variable S with values in $\mathcal{S}$, we denote by W_S the channel $\mathcal{X} \to \mathcal{Y}$ defined by

$$W_S(\cdot \mid \cdot) = EW(\cdot \mid \cdot, S). \tag{3.1}$$

For any DMC $\{W\}$, we denote by $C(W,\Gamma)$ its capacity under (peak) input constraint Γ, that is,

$$C(W,\Gamma) \stackrel{\Delta}{=} \max_{X:Eg(X)\le\Gamma, P_{Y|X}=W} I(X \wedge Y), \tag{3.2}$$

and define for the AVC $\{\mathcal{W}\}$

$$C(\Gamma,\Lambda) \stackrel{\Delta}{=} \min_{S:E\ell(S)\le\Lambda} C(W_S,\Gamma). \tag{3.3}$$

Lemma 3.1 For every $\Gamma > 0, \Lambda > 0$,

$$\begin{aligned} C(\Gamma,\Lambda) &= \min_{S:E\ell(S)\le\Lambda} \max_{X:Eg(X)\le\Gamma} I(X \wedge Y_{X,S}) \\ &= \max_{X:Eg(X)\le\Gamma} \min_{S:E\ell(S)\le\Lambda} I(X \wedge Y_{X,S}) \end{aligned} \tag{3.4}$$

where X and S denote independent r.v.s and Y is a random variable such that $Pr\{Y = y \mid X = x, S = s\} = W(y \mid x, s)$, or, $P_{Y|X} = W_S$. Furthermore, $C(\Gamma,\Lambda)$ is a nondecreasing, continuous, concave function of Γ, and a nonincreasing, continuous, convex function of Λ.

Proof: $I(X \wedge Y_{X,S})$ is convex in P_S since I is convex in $P_{Y|X} = W_S$ which, in turn, is linear in P_S (by eq. (3.1)); also $I(X \wedge Y_{X,S})$ is concave in P_X. Since $\{P_S : E\ell(S) \le \Lambda\}$ and $\{P_X : Eg(X) \le \Gamma\}$ are compact convex sets, the equality in (3.4) follows from the Minimax theorem (cf. e.g., Karlin [21]).

The convexity of $I(X \wedge Y_{X,S})$ in P_X implies in a standard manner that $\min_{S:E\ell(S)\leq\Lambda} I(X \wedge Y_{X,S})$ is a convex function of Λ. Thus $C(\Gamma,\Lambda)$ is the maximum of a family of convex functions of Λ and, hence, is itself convex. The concavity of $C(\Gamma,\Lambda)$ as a function of Γ follows similarly. The nondecreasing (resp. nonincreasing) property is trivial, and the continuity follows from the concavity (resp. convexity) property.

Lemma 3.2 For any S with $E\ell(S) < \Lambda$, and any $\epsilon' > \epsilon > 0$, every (n,Λ,ϵ)-random code for the AVC $\{W\}$ satisfies

$$E\bar{e}(W_S^n, F, \Phi) < \epsilon' \tag{3.5}$$

for the DMC $\{W_S\}$ defined by eq. (3.1), if n is large enough.

Proof: Let $\mathbf{S} = (S_1, ..., S_n)$ be n independent and identically distributed repetitions of a random variable S with $E\ell(S) < \Lambda$. Then, for every $\mathbf{x}$ in $\mathcal{X}^n$, $\mathbf{y}$ in $\mathcal{Y}^n$, by (2.2) and (3.1), we have

$$\begin{aligned} W_S^n(\mathbf{y} \mid \mathbf{x}) &= \prod_{i=1}^{n} EW(y_i \mid x_i, S_i) \\ &= E\prod_{i=1}^{n} W(y_i \mid x_i, S_i) \\ &= EW^n(\mathbf{y} \mid \mathbf{x}, \mathbf{S}). \end{aligned} \tag{3.6}$$

Any code (f,ϕ) when used on the memoryless channel $\{W_S\}$ defined by eq. (3.1) gives

$$\begin{aligned} \bar{e}(W_S^n, f, \phi) &= \frac{1}{\mid \mathcal{M} \mid} \sum_{m\epsilon\mathcal{M}} [1 - W_S^n(\phi^{-1}(m) \mid f(m))] \\ &= \frac{1}{\mid \mathcal{M} \mid} \sum_{m\epsilon\mathcal{M}} [1 - EW^n(\phi^{-1}(m) \mid f(m), \mathbf{S})] \text{ (by(3.6))} \\ &= E\bar{e}(\mathbf{S}, f, \phi) \\ &\leq E[\bar{e}(\mathbf{S}, f, \phi) \mid \ell(\mathbf{S}) \leq \Lambda] + Pr\{\ell(\mathbf{S}) > \Lambda\}, \end{aligned}$$

which implies that any random code (F,Φ) used on the DMC $\{W_S\}$ satisfies

$$E\bar{e}(W_S^n, F, \Phi) \leq E[\bar{e}(\mathbf{S}, F, \Phi) \mid \ell(\mathbf{S}) \leq \Lambda] + Pr\{\ell(\mathbf{S}) > \Lambda\}. \tag{3.7}$$

If (F, Φ) is an (n, Λ, ϵ)-random code for the AVC $\{\mathcal{W}\}$, the first term on the right side of (3.7) is clearly no larger than ϵ. Also, for n large enough, using the independent and identically distributed property of the S_i and the weak law of large numbers, the second term on the right side of (3.7) will be less than $(\epsilon' - \epsilon)$. Thus (3.7) gives (3.5).

We now state the random coding theorem for the case of peak constraints on the input and state sequences.

Note: Theorems 3.1 and 3.2 below (without the input constraints) were announced in [19] without proofs. The proofs presented here are new.

Theorem 3.1 For the AVC $\{\mathcal{W}\}$ with peak input/peak state constraints, the strong capacity (for random codes) exists and equals $C(\Gamma, \Lambda)$ defined by eq. (3.3).

Proof: The proof is similar to the case of the AVC without constraints (cf. e.g., [12, chapter 2, section 6]). The forward part of the proof is relegated to the appendix; the (strong) converse part is proved below.

We first observe that

$$C(\Gamma, \Lambda) = \inf_{S: E\ell(S) < \Lambda} C(W_S, \Gamma). \tag{3.8}$$

In fact, the right side can be written as a double infimum, the inner one for S with $E\ell(S) \leq \Lambda'$ and the outer one for $\Lambda' < \Lambda$. The inner infimum equals $C(\Gamma, \Lambda')$ by definition, and (3.8) follows by the monotonicity and continuity of $C(\Gamma, \Lambda)$ as a function of Λ.

Now, for any given $R > C(\Gamma, \Lambda)$ and $0 < \epsilon < 1$, pick an ϵ' with $\epsilon < \epsilon' < 1$ and a random variable S with $E\ell(S) < \Lambda$ such that

$$R > C(W_S, \Gamma),$$

which is possible by (3.8). Then, by the strong converse to the coding theorem for the DMC $\{W_S\}$ with (peak) input constraint Γ, every code (f, ϕ) of rate $\geq R$ satisfying the input constraint $g(f(m)) \leq \Gamma$ for all $m \epsilon \mathcal{M}$, has an average probability of error $\bar{e}(W_S^n, f, \phi) \geq \epsilon'$ if

$n \geq n_0$. Thus, for every random code (F, Φ) of rate $\geq R$ such that $g(F(m)) \leq \Gamma$ a.s., for all $m \epsilon M$, we have

$$E\bar{e}(W_S^n, F, \Phi) \geq \epsilon'.$$

This implies, by Lemma 3.2, that no such (F, Φ) can be an (n, Λ, ϵ)-code for the AVC $\{\mathcal{W}\}$, if n is sufficiently large. This proves that no rate above $C(\Gamma, \Lambda)$ is ϵ-achievable, for any $0 < \epsilon < 1$.

For the remaining combinations of peak and average constraints on the input and state sequences, the ϵ-capacities do depend on ϵ.

Theorem 3.2 For the AVC $\{\mathcal{W}\}$ with peak input, average state constraints the ϵ-capacity $C_\epsilon(\Gamma, a\text{-}\Lambda)$ equals $C(\Gamma, \Lambda/\epsilon)$, defined by eq. (3.3). In particular, the (weak) capacity equals $C(\Gamma, \ell_{\max})$ where $\ell_{\max} = \max_{s \epsilon S} \ell(s)$.

Proof: (i) $C_\epsilon(\Gamma, a\text{-}\Lambda) \geq C(\Gamma, \Lambda/\epsilon)$ (forward part):

To show that $C(\Gamma, \Lambda/\epsilon)$ is ϵ-achievable, on account of the continuity of $C(\Gamma, \Lambda)$ in Λ, it sufficies to show that any $R < C(\Gamma, \Lambda/\epsilon')$ is ϵ-achievable whenever $0 < \epsilon' < \epsilon$.

Theorem 3.1 implies that if $R < C(\Gamma, \Lambda/\epsilon')$, for n large enough, there exists a random code (F, Φ) of rate R with $g(F(m)) \leq \Gamma$ a.s. for all m in $\mathcal{M}$, that satisfies

$$E\bar{e}(\mathbf{s}, F, \Phi) \leq \epsilon - \epsilon' \text{ for all } \mathbf{s} \text{ with } \ell(\mathbf{s}) \leq \frac{\Lambda}{\epsilon'}.$$

Hence, for any random state sequence $\mathbf{S}$ with $E\ell(\mathbf{S}) \leq \Lambda$, we have

$$\begin{aligned} E\bar{e}(\mathbf{S}, F, \Phi) &= E[\bar{e}(\mathbf{S}, F, \Phi)/\ell(\mathbf{S}) \leq \frac{\Lambda}{\epsilon'}]\ Pr\{\ell(\mathbf{S}) \leq \frac{\Lambda}{\epsilon'}\} \\ &\quad + E[\bar{e}(\mathbf{S}, F, \Phi)/\ell(\mathbf{S}) > \frac{\Lambda}{\epsilon'}]\ Pr\{\ell(\mathbf{S}) > \frac{\Lambda}{\epsilon'}\} \\ &\leq E[\bar{e}(\mathbf{S}, F, \Phi)/\ell(\mathbf{S}) \leq \frac{\Lambda}{\epsilon'}] + Pr\{\ell(\mathbf{S}) > \frac{\Lambda}{\epsilon'}\} \\ &\leq (\epsilon - \epsilon') + \epsilon' = \epsilon. \end{aligned}$$

This means that (F, Φ) is an $(n, a\text{-}\Lambda, \epsilon)$-random code, and part (i) is proved.

(ii) $C_\epsilon(\Gamma, a\text{-}\Lambda) \leq C(\Gamma, \Lambda/\epsilon)$ (Converse part):

To show that no $R > C(\Gamma, \Lambda/\epsilon)$ is ϵ-achievable, it suffices to show this for $R > C(\Gamma, \Lambda/\epsilon')$, whenever $\epsilon' > \epsilon$. Pick $R > C(\Gamma, \Lambda/\epsilon')$ and let (F, Φ) be any random code of rate R satisfying the peak input constraint Γ. Theorem 3.1 implies that for sufficiently large n, the average error probability under peak state constraint Λ/ϵ' cannot be smaller than any fixed $\eta < 1$. This means that

$$E\bar{e}(\mathbf{s}, F, \Phi) \geq \eta$$

for some $\mathbf{s} = (s_1, ..., s_n)$ with $\ell(\mathbf{s}) \leq \Lambda/\epsilon'$. Now, let $\mathbf{S} = (S_1, ..., S_n)$ be a random state sequence such that $\mathbf{S} = \mathbf{s}$ with probability ϵ' and $\ell(\mathbf{S}) = 0$ with probability $(1 - \epsilon')$. Then, $E\ell(\mathbf{S}) \leq \Lambda$, and

$$\begin{aligned} E\bar{e}(\mathbf{S}, F, \Phi) &\geq \epsilon' E[\bar{e}(\mathbf{S}, F, \Phi) \mid \mathbf{S} = \mathbf{s}] \\ &= \epsilon' E\bar{e}(\mathbf{s}, F, \Phi) \geq \epsilon'\eta. \end{aligned}$$

Choosing $\eta = \epsilon/\epsilon'$, it follows that no random code of rate $R > C(\Gamma, \Lambda/\epsilon')$, satisfying the peak input constraint Γ, can be an $(n, a\text{-}\Lambda, \epsilon)$-random code. This proves part (ii).

The last assertion follows as $\lim_{\epsilon \to 0} C(\Gamma, \Lambda/\epsilon) = C(\Gamma, \ell_{\max})$.

Next, we prove a lemma and its corollary for a DMC with m-average and cm-average input constraints. These will be used in establishing the converse parts of Theorem 3.3.

Lemma 3.3 For any DMC $\{W\}$, any $\delta > 0$, and

$$R \geq C(W, \Gamma) + \delta, \tag{3.9}$$

(cf. eq. (3.2)), every code (f, ϕ) of blocklength $n \geq n_0$ and rate R has average error probability

$$\bar{e}(W^n, f, \phi) \geq 1 - \delta - \frac{\bar{g}(f)}{\Gamma} \tag{3.10}$$

where n_0 depends only on δ and the alphabet size $|\mathcal{X}|, |\mathcal{Y}|$.

Proof: We assume that $\bar{g}(f) < (1 - \delta)\Gamma$ in order to avoid a trivial assertion. Partition the message set $\mathcal{M}$ as $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ with $\mathcal{M}_1 = \{m : g(f(m)) \leq \Gamma\}$ and $\mathcal{M}_2 = \{m : g(f(m)) > \Gamma\}$. If $|\mathcal{M}_1| = \alpha |\mathcal{M}|$ and $|\mathcal{M}_2| = (1 - \alpha) |\mathcal{M}|$, $0 < \alpha < 1$, then clearly $(1 - \alpha)\Gamma \leq \bar{g}(f)$, i.e.,

$$\alpha \geq 1 - \frac{\overline{g}(f)}{\Gamma} > \delta. \tag{3.11}$$

Now consider the subcode of (f, ϕ) with message set M_1 and satisfying the (peak) input constraint Γ. The rate of this subcode is $R + \frac{1}{n}\log\alpha > R + \frac{1}{n}\log\delta > C(W,\Gamma) + \frac{\delta}{2}$ if $n \geq n_0$. Hence, by the strong converse to the coding theorem for a DMC (specifically by [12, p. 104, corollary 1.4]), the average error probability of this subcode is at least $(1-\delta)$ if $n \geq n_0$ (depending only on $\delta, |\mathcal{X}|, |\mathcal{Y}|$).

Thus, using (3.11), we have

$$\begin{aligned}
\overline{e}(W^n, f, \phi) &= \frac{1}{|M|}\sum_{m \in M} e_m(W^n, f, \phi) \\
&\geq \frac{\alpha}{|M_1|}\sum_{m \in M_1} e_m(W^n, f, \phi) \\
&\geq \alpha(1-\delta) \geq (1 - \frac{\overline{g}(f)}{\Gamma})(1-\delta) \\
&\geq 1 - \delta - \frac{\overline{g}(f)}{\Gamma}.
\end{aligned}$$

Corollary: For a DMC $\{W\}$, any random code (F, Φ) of rate $R > C(W,\Gamma) + \delta$ and blocklength $n \geq n_0$ (as in Lemma 3.3) has

$$E\overline{e}(W^n, F, \Phi) \geq 1 - \delta - \frac{E\overline{g}(F)}{\Gamma}. \tag{3.12}$$

Proof: First observe that Lemma 3.3 immediately implies (3.12) for random codes (F, Φ) for which $\overline{g}(F)$ is constant. In fact, were (3.12) not to hold in this case, there would exist a realization (f, ϕ) (a deterministic code) of the random code (F, Φ) violating (3.10). It follows then for an arbitrary (F, Φ) of rate satisfying (3.9) that

$$E[\overline{e}(W^n, F, \Phi) \mid \overline{g}(F)] \geq 1 - \delta - \frac{\overline{g}(F)}{\Gamma},$$

and hence

$$E\bar{e}(W^n, F, \Phi) = E[E[\bar{e}(W^n, F, \phi) \mid \bar{g}(F))]]$$
$$\geq 1 - \delta - \frac{E\bar{g}(F)}{\Gamma},$$

as claimed.

Theorem 3.3 For the AVC $\{\mathcal{W}\}$ with m-average or cm-average input constraints and peak state constraint, the ϵ-capacities $C_\epsilon(m\text{-}\Gamma, \Lambda)$ and $C_\epsilon(cm\text{-}\Gamma, \Lambda)$ are the same and equal $C(\Gamma/(1-\epsilon), \Lambda)$, defined by eq. (3.3). In particular, the (weak) capacity for both cases equals $C(\Gamma, \Lambda)$.

Proof: In order to prove the theorem, we need only prove:

(i) the forward part for the m-average input constaint; and

(ii) the converse part for the cm-average input constraint.

(i) We show that any $R < C(\Gamma/(1\text{-}\epsilon'), \Lambda)$ is ϵ-achievable under m-average input/peak state constraints whenever $\epsilon' < \epsilon$. Partition the message set $\mathcal{M}$ as $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ such that $\frac{|\mathcal{M}_1|}{|\mathcal{M}|} = 1 - \epsilon_n$, with $\epsilon_n \to \epsilon'$. If $R < C(\Gamma/(1-\epsilon'), \Lambda)$, by Theorem 3.1 there exists, for n sufficiently large, an $(n, \Lambda, \frac{\epsilon-\epsilon'}{2})$ random code (F, Φ) with message set $\mathcal{M}_1$ and satisfying the input constraint $g(F(m)) \leq \frac{\Gamma}{1-\epsilon'}$ a.s., i.e.,

$$E[\frac{1}{|\mathcal{M}_1|} \sum_{m \epsilon \mathcal{M}_1} e_m(\mathbf{s}, F, \Phi)] \leq \frac{\epsilon - \epsilon'}{2} \text{ for all } \mathbf{s} \epsilon \mathcal{S}^n \text{ with } \ell(\mathbf{s}) \leq \Lambda.$$

Let (F', Φ') be a random code which equals (F, Φ) whenever $m \epsilon \mathcal{M}_1$, and maps each m in $\mathcal{M}_2$ into a constant sequence $(x_0, ..., x_0)$ with $g(x_0) = 0$. Then $\frac{1}{|\mathcal{M}|} \sum_{m \epsilon \mathcal{M}} g(F'(m)) \leq \Gamma$ a.s., and for any $\mathbf{s}$ in $\mathcal{S}^n$ with $\ell(\mathbf{s}) \leq \Lambda$, we have

$$E[\bar{e}(\mathbf{s}, F', \Phi')] = E\frac{1}{|\mathcal{M}|}[\sum_{m \epsilon \mathcal{M}_1} e_m(\mathbf{s}, F, \Phi) + \sum_{m \epsilon \mathcal{M}_2} e_m(\mathbf{s}, F', \Phi')]$$
$$\leq \frac{\epsilon - \epsilon'}{2} + \epsilon_n \leq \epsilon.$$

This proves part (i).

(ii) It suffices to show that no $R > C(\Gamma/(1-\epsilon'), \Lambda)$ is ϵ-achievable under the cm-average input/peak state constraints whenever $\epsilon' > \epsilon$.

Given any $R > C(\Gamma/(1-\epsilon'), \Lambda)$, there exists by (3.8) a random variable S with $E\ell(S) < \Lambda$ such that

$$R > C(W_S, \Gamma/(1-\epsilon')) + \delta$$

for some $\delta > 0$; we may assume that $\delta < \epsilon' - \epsilon$. Now, by the Corollary to Lemma 3.3, any random code (F, Φ) of rate $\geq R$ satisfying the cm-average input constraint $E\overline{g}(F) \leq \Gamma$, has for the DMC $\{W_S\}$

$$E\overline{e}(W_S^n, F, \Phi) \geq 1 - \delta - \frac{E\overline{g}(F)}{\Gamma/(1-\epsilon')} \geq \epsilon' - \delta$$

if $n \geq n_0$. Since $\epsilon' - \delta > \epsilon$, this implies by Lemma 3.2, for n sufficiently large, that this (F, Φ) cannot be an (n, Λ, ϵ)-random code for the AVC $\{W\}$, as claimed.

For the remaining case of average input/average peak constraints, we have not been able to determine the ϵ-capacity. However, the (weak) capacity is easily obtained from the previous results:

Theorem 3.4. For m-average (resp. cm-average) input constraint m-Γ (resp. cm-Γ) and average state constraint a-Λ the (weak) capacity of the AVC $\{W\}$ equals $C(\Gamma, \ell_{\max})$.

Proof. The forward part immediately follows from Theorem 3.2, as $C(\Gamma, \ell_{\max})$ is an achievable rate even under the peak input constraint Γ. The converse part follows from Theorem 3.3 as the peak state constraint with $\Lambda = \ell_{\max}$ is always fulfilled.

4. Discussion

Random coding techniques serve as useful mathematical tools for proving coding theorems for a conventional (fixed) channel. Their use is justified by the fact that if the expected value of the decoding error probability over an ensemble of randomly selected codes is small, then there must exist a specific (deterministic) code leading to an error probability just as small. Thus, for a fixed channel, the deterministic code capacity equals the random code capacity.

In sharp contrast, an AVC exhibits the characteristic that the capacity for random codes generally exceeds that for deterministic codes. Consequently, as Ericson [16] remarks, in addition to helping prove coding theorems, random codes become significant as models of practical engineering devices. In fact, commonly used techniques such as "direct sequence" and "frequency hopping" can be interpreted as practical implementations of random codes [17], employing synchronized random number generators at the transmitter and receiver. The practical feasibility of random codes for AVCs is greatly enhanced by Ahlswede's [4] discovery that the random code capacity of an AVC can be achieved by codes restricted to random selections from no more than n^2 deterministic codes. This results in a desirably drastic reduction in the amount of additional information needed to convey the result of the random experiment of code selection from the encoder to the decoder across a special channel; in the terminology of Ericson [15], the "key rate" may be arbitrarily small.

In this paper we have determined the ϵ-capacities of the AVC for random codes under various, though not all possible, combinations of input and state constraints. The strong capacity turned out to exist only in the case of peak input, peak state constraints. The weak capacity was determined for all possible combinations of input and state constraints. It is interesting to note that even the discrete memoryless channel does not have a strong converse under the m-average input constraint, and that the bound given in Lemma 3.3 is actually tight (up to replacing $-\delta$ by $+\delta$); this simple fact has apparently not been pointed out before in the literature.

We did not consider the problem of whether, and under what conditions, deterministic codes can achieve the same capacities as random codes. The elimination technique of Ahlswede [4] gives that Theorem 3.1 remains valid for random codes restricted to random selections out of no more than n^2 deterministic codes. The final step of the elimination of randomness in [4], which intuitively means using a small fraction of the codeword to inform the decoder of

which of the n^2 codes was actually used, however cannot be performed unless the capacity (for deterministic codes) is positive even without a state constraint. Hence the capacity problem for deterministic codes has remained open for many practically interesting models. In a forthcoming paper [14], we will determine the deterministic code capacity of the AVC with (peak) constraints on the transmitted codewords as well as on the state sequences, and demonstrate that it may be positive but less than the corresponding random code capacity.

The ϵ-capacities (resp. weak capacity) of the AVC given by Theorems 3.1-3.3 (resp. Theorem 3.4) remain unchanged if the input (resp. state) constraints are imposed on each individual symbol of a sequence of length n, rather than on the sequence itself. Under the stronger symbol constraints, this holds by dint of the choice of codeword (resp. state) sequences used in the proofs of the forward (resp. converse) parts of Theorems 3.1-3.3. Furthermore, the results in this paper can be easily extended to the case of several constraints imposed simultaneously on the input (resp. state) sequence. For example, suppose that the random state sequence $\mathbf{S}$ is required to satisfy both the average constraint $E\ell(\mathbf{S}) \leq \Lambda$ and the peak constraint $\ell(\mathbf{S}) \leq \Lambda'$ a.s., with $\Lambda' > \Lambda$. It then follows, just as in Theorem 3.2, that the ϵ-capacity under the peak input constraint Γ equals $C(\Gamma, \min\{\Lambda', \Lambda/\epsilon\})$.

Our results do not depend in an essential way on the assumption $| \mathcal{S} | < \infty$. Most of the arguments hold also for infinite channel input and output alphabets, and in particular for the Gaussian AVCs considered in [20]. We believe that the approach in this paper makes the results in [20] more transparent. One difficulty in the general nondiscrete alphabet case appears to be with Lemma A.2, where the analog of $L_n(\mathbf{X}, \mathbf{Y})$ may not have a finite variance. Of course, our results cannot be expected to hold, without additional hypotheses, for AVCs with infinite alphabets because not even the strong converse for a memoryless channel does (cf. [8]).

Appendix

Forward Part of Theorem 3.1

Following the proof in Csiszár and Körner [12, pages 211-214], we establish the ϵ-achievability of $C(\Gamma, \Lambda)$, for each $0 < \epsilon < 1$, by the following two lemmas and their corollaries.

Lemma A.1 Let X be a random variable satisfying $g(X) \leq \Gamma$ a.s., and let d be a non-negative valued function on $\mathcal{X} \times \mathcal{Y}$ such that

$$Ed(X, y) \leq 1 \text{ for all } y \epsilon \mathcal{Y}. \tag{1}$$

Then there exists a random code (F, Φ) of block-length 1 with $g(F(m)) \leq \Gamma$ a.s. for all $m \epsilon \mathcal{M}$ such that for every channel $W : \mathcal{X} \to \mathcal{Y}$, every $m \epsilon \mathcal{M}$, and $\epsilon > 0$ arbitrary

$$Ee_m(W, F, \Phi) \leq Pr\{d(X, Y) \leq \frac{|\mathcal{M}|}{\epsilon}\} + \epsilon \tag{2}$$

where Y is connected with X by the channel W, that is, $P_{Y|X} = W$.

The proof is identical to that of Lemma 6.9 in [12, p. 211].

Corollary: If X and d are as in Lemma A.1 but X does not necessarily satisfy $g(X) \leq \Gamma$ a.s., there exists a random code (F, Φ) of blocklength 1 with $g(F(m)) \leq \Gamma$ a.s. for all $m \epsilon \mathcal{M}$, such that for every channel $W : \mathcal{X} \to \mathcal{Y}$, every $m \epsilon \mathcal{M}$, and $\epsilon > 0$ arbitrary

$$Ee_m(W, F, \Phi) \leq \epsilon + Pr\{d(X, Y) \leq \frac{|\mathcal{M}|}{\epsilon Pr\{g(X) \leq \Gamma\}}\} / Pr\{g(X) \leq \Gamma\}. \tag{3}$$

Proof: Apply Lemma A.1 to a random variable X' in the role of X such that the distribution of X' equals the conditional distribution of X given that $g(X) \leq \Gamma$, and to $d'(x, y) = d(x, y) \cdot Pr\{g(X) \leq \Gamma\}$ in the role of d. Then (1) holds since

$$Ed'(X', y) = Pr\{g(X) \leq \Gamma\} \cdot E[d(X, y) \mid g(X) \leq \Gamma] \leq Ed(X, y) \leq 1,$$

and (2) gives (3) because

$$Pr\{d'(X',Y') \le \frac{|\mathcal{M}|}{\epsilon}\} = Pr\{d(X,Y) \le \frac{|\mathcal{M}|}{\epsilon Pr\{g(X) \le \Gamma\}} / g(X) \le \Gamma\}$$

$$\le Pr\{d(X,Y) \le \frac{|\mathcal{M}|}{\epsilon Pr\{g(X) \le \Gamma\}}\} / Pr\{g(X) \le \Gamma\}.$$

For any distribution P on $\mathcal{X}$, and any channel $W : \mathcal{X} \to \mathcal{Y}$, let us denote by $I(P,W)$ the mutual information $I(X \wedge Y)$ for random variables X and Y connected by the channel W such that $P_X = P$.

Lemma A.2. For any r.v. with distribution $P_X = P$ such that $Eg(X) < \Gamma$, and any $\epsilon > 0$, for sufficiently large n there exist (n, Λ, ϵ)-random codes (F, Φ) for the AVC $\{\mathcal{W}\}$, satisfying $g(F(m)) \le \Gamma$ a.s. for all m in $\mathcal{M}$, and of rate at least $I(P,\Lambda) - \epsilon$ where

$$I(P,\Lambda) \triangleq \min_{S: E\ell(S) \le \Lambda} I(P, W_S). \tag{4}$$

Proof: Let W_{S_0} minimize $I(P,W_S)$ subject to $E\ell(S) \le \Lambda$. Then for every S such that $E\ell(S) \le \Lambda$, and for $0 \le \alpha \le 1$, we obtain by using the convexity of $\tilde{\mathcal{W}} \triangleq \{W_S : E\ell(S) \le \Lambda\}$, that

$$I(P, \alpha W_S + (1-\alpha) W_{S_0}) \ge I(P, W_{S_0}) = I(P,\Lambda). \tag{5}$$

Hence, it follows as in [12, page 213] that $P(x)W_S(y \mid x) > 0$ implies $W_{S_0}(y \mid x) > 0$, and that

$$\sum_{x,y} P(x) W_S(y \mid x) \log \frac{W_{S_0}(y \mid x)}{Q(y)} \ge I(P,\Lambda) \tag{6}$$

if $E\ell(S) \le \Lambda$, where $Q(y) \triangleq \sum_x P(X) W_{S_0}(y \mid x)$.

Let $\mathbf{X} = (X_1, ..., X_n)$ be a sequence of independent and identically distributed random variables with distribution P. We apply the Corollary to Lemma A.1 to $\mathcal{X}^n$ and $\mathcal{Y}^n$ in the roles of $\mathcal{X}$ and $\mathcal{Y}$, with this $\mathbf{X}$, and $d(\mathbf{x},\mathbf{y})$ defined by

$$d(\mathbf{x},\mathbf{y}) = \frac{W_{S_0}^n(\mathbf{y} \mid \mathbf{x})}{Q^n(\mathbf{y})} = \prod_{i=1}^n \frac{W_{S_0}(y_i \mid x_i)}{Q(y_i)} \tag{7}$$

and set $d(\mathbf{x},\mathbf{y}) = 1$ if $Q^n(\mathbf{y}) = 0$. Clearly, $Ed(\mathbf{X},\mathbf{y}) = 1$ for all $\mathbf{y}\epsilon\mathcal{Y}^n$. Then the said corollary guarantees the existence of a block-length n random code (F,Φ) such that for every $m\epsilon\mathcal{M}$, $g(F(m)) \leq \Gamma$ a.s., and for every $\mathbf{s}\epsilon\mathcal{S}^n$,

$$Ee_m(\mathbf{s},F,\Phi) \leq \frac{\epsilon}{2} + \frac{Pr\{\frac{W_{S_0}^n(\mathbf{Y}|\mathbf{X})}{Q^n(\mathbf{Y})} < \frac{2}{\epsilon}\frac{|\mathcal{M}|}{Pr\{g(\mathbf{X})\leq\Gamma\}}\}}{Pr\{g(\mathbf{X}) \leq \Gamma\}} \tag{8}$$

where $\mathbf{Y} = (Y_1,...,Y_n)$ satisfies $P_{\mathbf{Y}|\mathbf{X}}(\cdot \mid \cdot) = W^n(\cdot \mid \cdot,\mathbf{s})$.

Defining the random variable $L_n(\mathbf{X},\mathbf{Y})$ by

$$L_n(\mathbf{X},\mathbf{Y}) \triangleq \log \frac{W_{S_0}^n(\mathbf{Y} \mid \mathbf{X})}{Q^n(\mathbf{Y})}, \tag{9}$$

we observe that

$$\begin{aligned} EL_n(\mathbf{X},\mathbf{Y}) &= \sum_{\mathbf{x},\mathbf{y}} P^n(\mathbf{x})W^n(\mathbf{y} \mid \mathbf{x},\mathbf{s}) \log[\prod_{i=1}^n \frac{W_{S_0}(y_i \mid x_i)}{Q(y_i)}] \\ &= \sum_{i=1}^n \sum_{x,y} P(x)W(y \mid x,s_i) \log \frac{W_{S_0}(y \mid x)}{Q(y)}. \end{aligned} \tag{10}$$

Now, if $\mathbf{s}\epsilon\mathcal{S}^n$ satisfies $\ell(\mathbf{s}) \leq \Lambda$, then letting $\overline{\mathrm{S}}$ denote a random variable whose distribution equals the type of $\mathbf{s} = (s_1,...,s_n)$, we have

$$\sum_{i=1}^n W(y \mid x,s_i) = nW_{\overline{\mathrm{S}}}(y \mid x) \text{ with } E\ell(\overline{\mathrm{S}}) \leq \Lambda.$$

Thus, by (6), we obtain from (10) that

$$EL_n(\mathbf{X},\mathbf{Y}) \geq nI(P,\Lambda) \text{ if } \ell(\mathbf{s}) \leq \Lambda. \tag{11}$$

Furthermore, since $W_{S_0}(y \mid x) > 0$ whenever $P(x)W_S(y \mid x) > 0$, it can be seen as in [12, page 214] that

$$\text{var } L_n(\mathbf{X},\mathbf{Y}) \leq n[\log m(W_{S_0})]^2, \tag{12}$$

where $m(W_{S_0})$ is the smallest positive entry in W_{S_0}.

Since $Eg(X) < \Gamma$, for sufficiently large n we have $Pr\{g(\mathbf{X}) \leq \Gamma\} > 1-\epsilon$. Then if

$$\mid \mathcal{M} \mid = \exp[n\{I(P,\Lambda) - \epsilon\}], \tag{13}$$

from eqs. (8) - (13), we obtain by Chebyshev's inequality for every $\mathbf{s}$ in $\mathcal{S}^n$ satisfying $\ell(\mathbf{s}) \leq \Lambda$ that

$$\begin{aligned} Ee_m(\mathbf{s},F,\Phi) &\leq \frac{\epsilon}{2} + \frac{Pr\{L_n(\mathbf{X},\mathbf{Y}) < \log \frac{|\mathcal{M}|}{\epsilon(1-\epsilon)/2}\}}{(1-\epsilon)} \\ &\leq \frac{\epsilon}{2} + \frac{Pr\{L_n(\mathbf{X},\mathbf{Y}) < n(I(P,\Lambda)-\epsilon) - \log \frac{\epsilon(1-\epsilon)}{2}\}}{(1-\epsilon)} \\ &\leq \frac{\epsilon}{2} + \frac{Pr\{\mid L_n - EL_n \mid > \frac{n\epsilon}{2}\}}{(1-\epsilon)} \\ &\leq \frac{\epsilon}{2} + \frac{4}{n\epsilon^2(1-\epsilon)} \log[m(W_{S_0})]^2. \end{aligned} \tag{14}$$

It follows from (14) that for n sufficiently large

$$Ee_m(\mathbf{s},F,\Phi) \leq \epsilon \text{ if } \ell(\mathbf{s}) \leq \Lambda$$

for all m in $\mathcal{M}$; thus (F,Φ) is an (n,Λ,ϵ) random code. This proves the assertion of the lemma.

Corollary: $C(\Gamma,\Lambda)$ is an achievable rate.

Proof: It follows from Lemma A.2 that there exist (n,Λ,ϵ) random block codes (F,Φ) with $g(F(m)) \leq \Gamma$ a.s. for all m in $\mathcal{M}$, having rates arbitrarily close to

$$\sup_{X:Eg(X)<\Gamma} I(P_X,\Lambda).$$

By Lemma 3.1, it can be seen by analogy with (3.8) that here the $<$ sign can be replaced by $\leq$, and that the supremum equals $C(\Gamma,\Lambda)$. This completes the proof.

References

[1] R. Ahlswede, "The Weak Capacity of Averaged Channels", Z. *Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 11, pp. 61-73, 1968.

[2] R. Ahlswede, "A Note on the Existence of the Weak Capacity for Channels with Arbitrarily Varying Channel Probability Functions and its Relation to Shannon's Zero Error Capacity," *Ann. Math. Statist.*, vol. 41, pp. 1027-1033, 1970.

[3] R. Ahlswede, "The Capacity of a Channel with Arbitrarily Varying Additive Gaussian Channel Probability Functions", *Trans. Sixth Prague Conf. on Inform. Theory, Stat. Dec. Functions and Random Process*, pp. 13-21, 1971.

[4] R. Ahlswede, "Elimination of Correlation in Random Codes for Arbitrarily Varying Channels," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 44, pp. 159-175, 1978.

[5] R. Ahlswede, "A Method of Coding and an Application to Arbitrarily Varying Channels," *J. Combinatorics, Inform. Systems Sciences*, vol. 5, pp. 10-35, 1980.

[6] R. Ahlswede, "Arbitrarily Varying Channels with States Sequence Known to the Sender," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 621-629, September 1986.

[7] R. Ahlswede and J. Wolfowitz, "The Capacity of a Channel with Arbitrarily Varying Channel Probability Functions and Binary Output Alphabet," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 15, pp. 186-194, 1970.

[8] U. Augustin, "Gedächtnisfreie Kanäle für Diskrete Zeit," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 6, pp. 10-61, 1967.

[9] N.M. Blachman, "On the Capacity of a Band-Limited Channel Perturbed by Statistically Dependent Interference", *IRE Trans. Inform. Theory*, vol. IT-8, pp. 48-55, January 1962.

[10] N.M. Blachman, "The Effect of Statistically Dependent Interference upon Channel Capacity," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 553-557, September 1962.

[11] D. Blackwell, L. Breiman, and A.J. Thomasian, "The Capacities of Certain Channel Classes under Random Coding," Ann. Math. Stat., vol. 31, pp. 558-567, 1960.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[13] I. Csiszár and J. Körner, "On the Capacity of the Arbitrarily Varying Channel for Maximum Probability of Error," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 57, pp. 87-101, 1981.

[14] I. Csiszár and P. Narayan, "The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, State Constraints," submitted to the *IEEE Trans. Inform. Theory.*

[15] T. Ericson, "Exponential Error Bounds for Random Codes in the Arbitrarily Varying Channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 42-48, January 1985.

[16] T. Ericson, "The Non-Cooperative Binary Adder Channel," *IEEE Trans. Inform. Theory*, vol IT-32, pp. 365-374, May 1986.

[17] T. Ericson, "A Min-Max Theorem for Antijamming Group Codes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 792-799, November 1984.

[18] S.I. Gel'fand and M.S. Pinsker, "Coding for Channels with Random Parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19-31, 1980.

[19] B. Hughes and P. Narayan, "Interleaving and Channels with Unknown Memory," *Proc. Conf. Inform. Sciences and Systems*, The Johns Hopkins University, Baltimore, Maryland, February 1985.

[20] B. Hughes and P. Narayan, "Gaussian Arbitrarily Varying Channels," *IEEE Trans. Inform. Theory*, to appear, March 1987.

[21] S. Karlin, *Mathematical Methods and Theory in Game, Programming and Economics*, Addison-Wesley, Reading, MA, 1959.

[22] J. Kiefer and J. Wolfowitz, "Channels with Arbitrarily Varying Channel Probability Functions," *Inform. Contr.*, vol. 5, pp. 44-54, 1962.

[23] C.E. Shannon, "The Zero Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8-19, January 1956.

[24] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed., Springer-Verlag, Berlin, 1978.